

All Together Now - The Convergence of Big Data, Cloud, and HPC: A perspective from a new cloud provider



A national science & engineering cloud

Craig Stewart

ORCID ID 0000-0003-2423-9019

Jetstream Principal Investigator

Executive Director, Indiana University Pervasive Technology Institute

Wednesday, Nov 18, 1:30 p.m. – 2:30 p.m.

Dell Sixth Street Theater, booth #1009

funded by the National Science Foundation

Award #ACI-1445604

The issue in perspective...

- More diversity and open questions in processors than there has been in a long, long time
- More demand for a greater variety of computational resources than ever
- Overall, the economics of power overwhelming from cloud standpoints
- But customer 'capture' is an issue
- Universities, colleges, and companies cannot be effective and shirk financial responsibility for their cyberinfrastructure resources
- But it's very hard for any institution to invest in and manage resources at proper scale when you need low latency HPC, traditional clusters, support for Map – reduce based algorithms.
- What's an organization to do?
- Invest locally and globally; use locally and globally
- IU, TACC, University of Arizona, University of Chicago, and Johns Hopkins University are standing up a new production cloud resource with NSF funding



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

What is Jetstream and what will it do?

- Jetstream is
 - A First-Of-A-Kind NSF's first cloud for science and engineering research across all areas of activity supported by the NSF
 - designed to give researchers and research students access to interactive computing and data analysis resources “on demand.”
 - In the atmosphere at the border of two different air masses; in NSF-funded cyberinfrastructure at border of the existing NSF-funded XD program and advanced cyberinfrastructure resources and users who have not previously used such NSF funded infrastructure before.
 - A cloud: CentOS, Openstack, KVM, Atmosphere
- Jetstream will enable researchers and research students to:
 - Use Virtual Machines interactively
 - Move data to and from Jetstream using Globus Transfer
 - Use virtual desktops.
 - Publish VMs with a DOI



**RESEARCH
TECHNOLOGIES**

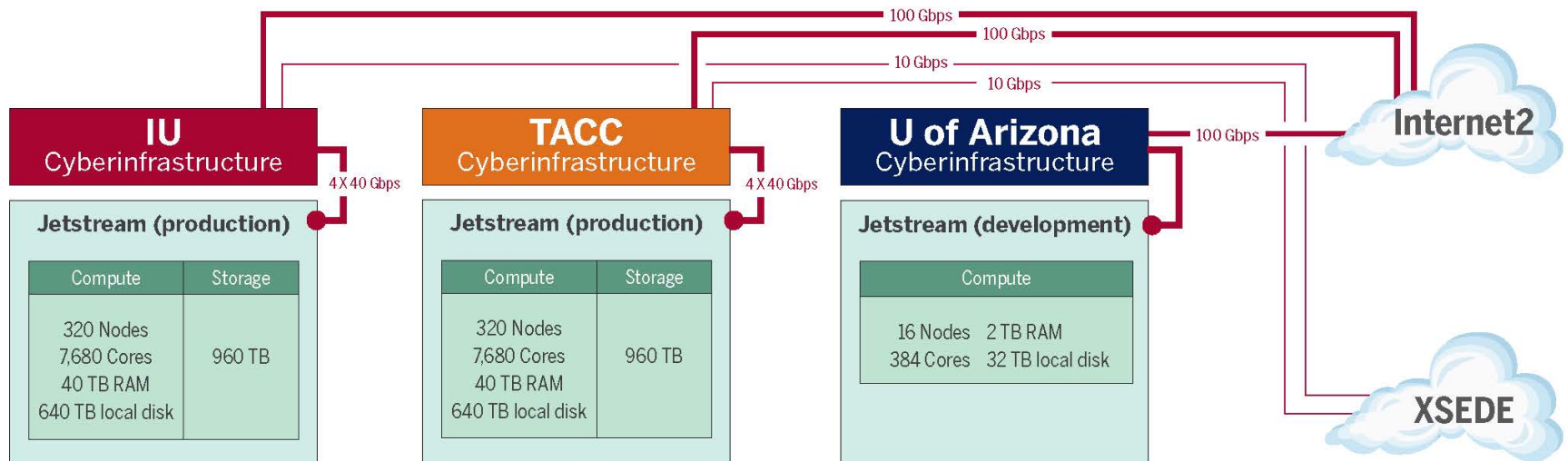
INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

Jetstream System Diagram



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY





DEMO

Dashboard

Projects

Images

Providers

Help

edwin@xsede.org

Getting Started



Launch New Instance

Browse Atmosphere's list of available images and select one to launch a new instance.



Browse Help Resources

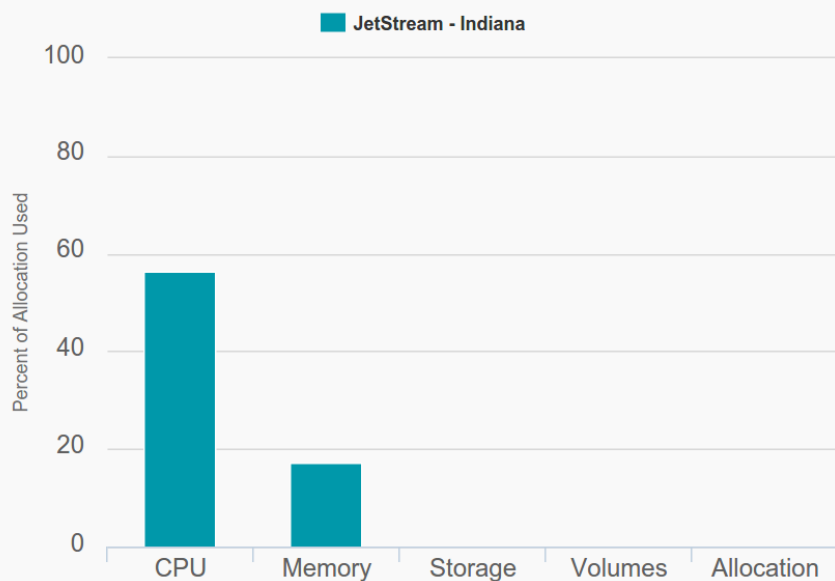
View a video tutorial, read the how-to guides, or email the Atmosphere support team.



Change Your Settings

Modify your account settings, view your resource quota, or request more resources.

Resources in Use [Need more?](#)



3 Instances

■ active - deploying



0 Volumes

Community Activity



atmoadmin created an image
Nov 05, 2015 12:23 pm
[MAKER-P_2.28_with_CCTools_5](#)



atmoadmin created an image
Nov 05, 2015 12:23 pm
[TSW_Workshop_Williams_1.2](#)



atmoadmin created an image
Oct 22, 2015 09:06 pm
[Trusty Tahr \(x64\)](#)



atmoadmin created an image
Oct 22, 2015 09:06 pm
[cirros-0.3.4-x86_64](#)



atmoadmin created an image
Oct 22, 2015 09:06 pm
[CentOS-7-x86_64-GenericCloud-20150628_01](#)



atmoadmin created an image
Oct 22, 2015 09:06 pm
[CentOS-6-x86_64-GenericCloud-1508](#)



atmoadmin created an image
Oct 22, 2015 09:06 pm
[CentOS-7-x86_64-GenericCloud-1508](#)

Jetstream-Indiana



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Status highlights

- Jetstream-Texas is HERE AT TACC! You can see Jetstream if you take the tour out to TACC!
- Jetstream was oversubscribed in its first round of several XRAC allocations
- Several startup allocations already being processed
- We have very limited openings for early adopters to start using Jetstream now
- Early operations mode is slated to start 20 January 2016
- NSF DGA (Division of Grants and Awards) has:
 - Declared acceptance of the text portion of the cluster (Jetstream-Arizona)
 - Has accepted our PEP (Program Execution Plan)
- For resources allocated through XSEDE fill out application form at <https://www.xsede.org/allocations>.



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

A few closing thoughts

- In terms of federally funded research there simply is not enough money in the system to do all we would all like to have done
- It's important to focus on commonalities wherever possible (e.g. node level performance)
- We need to recognize differences where they exist, innovate in clouds, innovate in HPC, and strive to create federated (or at least nationally accessible) systems of resources that meet our collective needs to support research, innovation, and the translation of innovation into practical benefits in science, engineering, and society



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

Jetstream Collaborators

- University of Chicago - Globus
- Arizona University – iPlant
- Johns Hopkins University and Penn State University
- Cornell University –Ms. Susan Mehringer, Lead. Cornell® Virtual Workshops about Jetstream and applications running on jetstream.
- University of Arkansas at Pine Bluff – Dr. Jesse Walker, lead. cybersecurity education,, Minority Serving Education outreach
- University of Hawaii – Dr. Gwen Jacobs, lead. EPSCoR early adopter/user. Jacobs will chair Science Advisory Board
- National Snow and Ice Data Center (NSIDC) – Dr. Ron Weaver, lead. Data retrieval from NSIDC, application integration with ice sheet analysis applications
- University of North Carolina, Odum Center –Dr. Thomas Carsey , lead. Data retrieval from Dataverse Network
- National Center for Genome Analysis at Indiana University – providing genome analysis software. Includes TACC, PSC, and SDSC as partners
- **Dell – our vendor partner! (Thank you!!!!)**



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

Acknowledgments & Disclaimers

- Jetstream is supported by NSF award 1445604 (Craig Stewart, PI)
- XSEDE is supported by NSF award 1053575 (John Towns, UIUC, PI)
- This work was also supported by the Indiana University Pervasive Technology Institute, which was initiated with major funding from the Lilly Endowment, Inc.
- Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF) or other supporting organizations.



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

Citation and license terms

- Please cite as Stewart, C.A. 2015. All Together Now - The Convergence of Big Data, Cloud, and HPC: A perspective from a new cloud provider. <http://hdl.handle.net/2022/22447>
- Except where otherwise noted, by inclusion of a source url or some other note, the contents of this presentation are © by the Trustees of Indiana University. This content is released under the Creative Commons Attribution 3.0 Unported license (<http://creativecommons.org/licenses/by/3.0/>). This license includes the following terms: You are free to share – to copy, distribute and transmit the work and to remix – to adapt the work under the following conditions: attribution – you must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). For any reuse or distribution, you must make clear to others the license terms of this work.



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY